DOCUMENT RESUME

ED 465 007 UD 035 060

AUTHOR Jacob, Brian A.; Lefgren, Lars
TITLE Remedial Education and Student Achievement: A Regression-Discontinuity Analysis.
PUB DATE 2001-09-00
NOTE 58p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Academic Achievement; Academic Standards; Accountability; Elementary Secondary Education; *Grade Repetition; Remedial Programs; Social Promotion; Standardized Tests; *Summer Programs; Urban Schools
IDENTIFIERS Chicago Public Schools IL

ABSTRACT

This study used a regression discontinuity design to examine the causal effect of summer school and grade retention on student achievement. In 1996, the Chicago Public Schools instituted an accountability policy that tied summer school attendance and promotional decisions to performance on standardized tests, which created a highly non-linear relationship between current achievement and probability of attending summer school or being retained. This study used the exogenous variation generated by the decision rule to identify the impact of these remedial programs. It utilized Chicago Public School data on test scores, demographics, bilingual and special education status, and residential and school mobility. Students in grades 3-6 were followed for 5 years. Results showed that summer school increased academic achievement in reading and mathematics and that these positive effects remained substantial at least 2 years following completion of the program. In contrast to prior studies, retention had no negative consequences on the academic achievement of students retained in third grade and in fact may have increased short-term performance. The impact of retention on older students was mixed, with no impact on math and a negative effect on reading. (Contains 7 figures, 9 tables, and 34 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 465 007

# REMEDIAL EDUCATION AND STUDENT ACHIEVEMENT: A REGRESSION-DISCONTINUITY ANALYSIS*

September 2001

Brian A. Jacob
John F. Kennedy School of Government
Harvard University

Lars Lefgren
Economics Department
Brigham Young University

Abstract

As standards and accountability have become an increasingly prominent feature of the educational landscape, educators have relied more on remedial programs such as summer school and grade retention to help low-achieving students meet minimum academic standards. Yet the evidence on the effectiveness of such programs is mixed, and prior research suffers from selection bias. However, recent school reform efforts in Chicago provide an opportunity to examine the causal impact of these remedial education programs. In 1996, the Chicago Public Schools instituted an accountability policy that tied summer school and promotional decisions to performance on standardized tests, which resulted in a highly non-linear relationship between current achievement and the probability of attending summer school or being retained. Using a regression discontinuity design, we find that the net effect of these programs was to substantially increase academic achievement among third graders, but not sixth graders. In addition, contrary to conventional wisdom and prior research, we find that retention increases achievement for third grade students and has little effect on math achievement for sixth grade students.

* We would like to thank the Consortium on Chicago School Research and the Chicago Public Schools for providing the data used in this study. We are grateful to Anthony Bryk, Thomas DeLeire, Mark Duggan, Michael Greenstone, Steven Levitt, Helen Levy, Brigitte Madrian, Casey Mulligan, Kevin Murphy, Melissa Roderick, Mark Showalter, and seminar participants at the University of Chicago for helpful suggestions. All remaining errors are our own. Jacob can be contacted at the John F. Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge, MA 02138 and by email at brian_jacob@harvard.edu. Lefgren can be contacted at the Department of Economics, Brigham Young University, 130 Faculty Office Building, Provo, UT 84602-2363 and by e-mail at l-lefgren@byu.edu.

UD 035 060

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

B. Jacob

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

REMEDIAL EDUCATION AND STUDENT ACHIEVEMENT:
A REGRESSION DISCONTINUITY ANALSYSIS


Abstract

As standards and accountability have become an increasingly prominent feature of the educational landscape, educators have relied more on remedial programs such as summer school and grade retention to help low-achieving students meet minimum academic standards. Yet the evidence on the effectiveness of such programs is mixed, and prior research suffers from selection bias. However, recent school reform efforts in Chicago provide an opportunity to examine the causal impact of these remedial education programs. In 1996, the Chicago Public Schools instituted an accountability policy that tied summer school and promotional decisions to performance on standardized tests, which resulted in a highly non-linear relationship between current achievement and the probability of attending summer school or being retained. Using a regression discontinuity design, we find that the net effect of these programs was to substantially increase academic achievement among third graders, but not sixth graders. In addition, contrary to conventional wisdom and prior research, we find that retention increases achievement for third grade students and has little effect on math achievement for sixth grade students.

## I. Introduction

Education is one of the most important avenues through which governments can address concerns of economic growth and equity. Human capital plays a substantial role in the economic growth of nations (Topel 1999) and, in the past two decades, skill biased technical change has increased the returns to schooling, exacerbating wage inequality between the most and least educated members of our society (Katz and Murphy 1992). At the same time, cognitive ability has become an increasingly important determinant of labor market success in this country (Murnane et al. 1995).

Aware of the importance of education, economists have spent considerable effort examining what factors affect academic achievement. There is a large literature on the importance of financial resources in determining educational outcomes.[1] However, researchers have paid considerably less attention to remedial programs designed to improve the performance of low achieving students, including summer school and grade retention (Eide and Showalter *forthcoming*).

Such policies, however, have become increasingly popular in recent years. Sixteen states provide funding for districts that institute summer programs, require summer school attendance for students who do not meet academic expectations or require districts to offer summer school to low achieving students (ECS 2000). In the summer of 1999, New York City provided summer help for 70,000 students and Chicago required over 30,000 low-achieving students to attend summer classes. Other urban districts with summer programs include Houston, with 8,000 students enrolled; Boston, with 6,500; Denver, with 6,000; Los Angeles, with 139,000; and the

District of Columbia, with 30,000 (Pipho 1999). There is a growing interest in grade retention as well. Nineteen states explicitly tie student promotion to performance on a state or district assessment (ECS 2000). The largest school districts in the country, including New York City, Los Angeles, Chicago and Washington D.C., have recently implemented policies requiring students to repeat a grade when they do not demonstrate sufficient mastery of basic skills.

Despite their popularity, these practices—particularly grade retention—remain controversial. Prior research suggests that summer school has a substantial positive effect on student learning in the short-run, but there is less evidence regarding the sustainability of achievement gains made during the summer. In contrast, the majority of retention studies find that the practice of requiring students to repeat a grade decreases self-esteem, school adjustment, and academic achievement, and increases dropout rates. However, prior studies fail to account for the selection of students into these programs, thus potentially overstating the benefits of summer school and exaggerating the harm of retention.

In this paper, we use a regression discontinuity design to examine the causal effect of summer school and grade retention on student achievement. In 1996, the Chicago Public Schools (CPS) instituted an accountability policy that tied summer school attendance and promotional decisions to performance on standardized tests, which resulted in a highly non-linear relationship between current achievement and the probability of attending summer school or being retained.[2] We use the exogenous variation generated by the decision rule to identify the impact of these remedial programs.

---

[1] Hanushek (1996) and Hedges and Greenwald (1996) present contrasting views regarding the effectiveness of resources on student achievement. Many researchers have also focused on the effect of class size on student outcomes. These studies include works by Angrist and Lavy (1999), Krueger (1999), and Hoxby (2000).

[2] In prior work, we have examined the potential motivational effects of these requirements and found that the policy increased achievement, particularly among older students (Jacob 2001; Roderick et. al. 2000). In this analysis, we

We find that summer school increased academic achievement in reading and mathematics and that these positive effects remain substantial at least two years following the completion of the program. In contrast to prior studies, we find that retention has no negative consequences on the academic achievement of students retained in the third grade—indeed it appears that retention may actually increase performance in the short run. The impact of retention on older students is mixed, with no impact on math and a negative effect on reading.[3]

The remainder of this paper is organized as follows. Section 2 reviews the previous literature on summer school and grade retention. Section 3 provides background on the Chicago policy. Section 4 describes our data and Section 5 explains our empirical strategy. Section 6 presents findings on the net effect of summer school and retention. Section 7 presents findings on the separate effect of grade retention. Section 8 examines the independent effect of summer school. Section 9 discusses these findings and concludes.

## 2. Previous Literature on Summer School and Grade Retention

Both summer school and grade retention have a long history within American education, dating back to the introduction of mass public education in the mid-nineteenth century (Shepard and Smith 1989). Both practices have been widely implemented and have received considerable attention from researchers. However, prior studies do not adequately address the potential biases introduced by the non-random selection into summer school and retention.

In a detailed synthesis of 93 summer school evaluations, Cooper et. al. (2000) concluded that remedial summer programs increased achievement by roughly .25 standard deviations.

---

set aside the incentives associated with the policy and instead focus on the direct academic consequences of summer school and grade retention for those students who fail to meet the promotional standards.

[3] However, as we discuss below, the negative effects for sixth grade students may be due to differential test incentives faced by retained and promoted students.

However, even the most careful of these studies relied on comparisons between students who chose to attend summer school with those who chose not to attend. If the most motivated students (or those with the most motivated, supportive parents) attend summer school, then the estimated summer school treatment effect will be biased upward. In addition, these studies do not examine whether these benefits are sustained in subsequent years.

While less consistent than the summer school literature, studies of grade retention have generally found that repeating a grade has a negative impact on student outcomes.[4] In a survey of 47 empirical studies with a variety of academic achievement measures, Holmes (1989) found that retained students scored 0.19 to 0.31 standard deviations below comparable students who had not been retained. Moreover, a variety of studies have found that retention is associated with an increased likelihood of dropping out (Schulz, Toles et al. 1986; Rumberger 1987; Grissom and Shepard 1989; Fine 1991; Roderick 1994). However, selection issues cast doubt on these findings as well. In contrast to summer school, students are not generally given a choice whether to repeat a grade, but rather this decision is made by the teacher or school principal on the basis of unobservable characteristics (e.g., motivation, maturity, parental involvement, etc.). This suggests that OLS estimates of grade retention will be biased downward.

### 3. Background on Chicago's Social Promotion Policy

An accountability policy recently implemented in Chicago provides an opportunity to more carefully examine the impacts of these programs. In 1996-97, Chicago instituted a policy to end social promotion – the practice of passing students to the next grade regardless of their academic skills or school performance. Under the policy, students in third, sixth and eighth

[4] Several recent studies have found moderate, positive effects of retention (Karweit 1991; Pierson and Connell 1992; Alexander, Entwisle et al. 1994; Eide and Showalter *forthcoming*; Dworkin, Lorence et al. 1999).

grade are required to perform at predefined levels in both reading and mathematics in order to be promoted to the next grade. For example, third graders must obtain a minimum score of 2.8 grade equivalents in both reading and math achievement on the Iowa Test of Basic Skills (ITBS) in order to advance.[5] In 1997, the promotion standards for third, sixth and eighth grade were 2.8, 5.3, and 7.0 respectively, which roughly corresponded to the 20th percentile in the national achievement distribution.[6] Students who do not meet the standard in June are required to attend a six-week summer school program, after which they can retake the exams. Those who pass the August exams move on to the next grade. Students who again fail are required to repeat the grade.[7] Figure 1 provides a flowchart illustrating the treatments children receive based on their June and August test performance. The policy impacted a large proportion of elementary students. From 1997 to 1999, over 30,000 third graders and over 21,000 sixth graders attended a mandatory summer school program and roughly 10 to 20 percent of the eligible students were eventually held back.

There are several reasons to believe that the summer school and grade retention programs in Chicago might influence academic achievement. Classes in summer school were generally quite small, often with fewer than 15 students per class. Principals hand-picked teachers for summer school and the CPS provided a highly structured curriculum (including resource materials) that teachers were required to follow. Retention was intended to provide students additional time to master the skills at their current grade level. The CPS also provided schools with additional resources to meet the needs of retained students.

---

[5] Grade equivalents are normed so that a student at the 50th percentile in the nation scores at the eighth month of her current grade – i.e., an average third grader will score a 3.8.

[6] The CPS has raised the promotional cutoffs several times since 1997. The eighth grade cutoff was raised to 7.2 in 1998, 7.4 in 1999 and 7.7 in 2000. The sixth grade cutoff was raised to 5.5 in 2000.

## 4. Data

This study utilizes administrative data from the Chicago Public School system. Student records provide individual level information on test scores and student demographics (race, gender, age, guardian, and free lunch eligibility), bilingual and special education status, and residential and school mobility. Unique student identification numbers allow us to follow individual students throughout their tenure in the public school system. School level data provides demographic and school resource information, including the racial and socio-economic composition at the school. The outcome measure we use is student scores on the math and reading sections of the Iowa Test of Basic Skills (ITBS), a standardized multiple-choice exam administered annually to students in grades three to eight.[8]

The base sample for this study consists of the cohort of students who were in the third and sixth grade from the 1993-1994 school year to the 1998-1999 school years, a total of 402,924 observations.[9] We delete approximately 14 percent of cases that were missing demographic data or third (or sixth) grade test scores, leaving a sample of 346,909 students. Additionally, we drop 45,534 individuals who were not subject to the promotion policy because

---

[7] Students over the age of fifteen who were retained were placed in special "transition" centers.

[8] ITBS scores are typically reported in terms of grade equivalents, which are normed so that a student at the 50th percentile in the nation scores at the eighth month of her current grade – i.e., an average third grader will score a 3.8. Grade equivalents, however, present a number of well-know shortcomings for comparisons over time and across grade: (1) different forms of the exam are administered each year and can vary in difficulty; (2) grade equivalents are not a linear metric, so that a score of 5.3 on level 12 of the exam does not represent the same thing as a score of 5.3 at level 13; (3) grade equivalents are not linear within test level because the scale spreads out more at the extremes of the score distribution. To mitigate some of these concerns, we use an alternative outcome metric derived from an item-response model. This model assume that the probability that student $i$ answers questions $j$ correctly is a function of the student ability and the item difficulty. In practice, one estimates a simple logit model in which the outcome is whether or not student $i$ correctly answers question $j$. The explanatory variables include an indicator variable for each question and each student. The difficulty of the question is given by the coefficient on the appropriate indicator variable and the student ability is measured by the coefficient on the student indicator variable. The resulting metric is calibrated in terms of logits. By taking advantage of the common items across different forms and levels of the exam, these measures provide an effective way to compare students in different grade levels or taking different forms of the exam (Wright and Stone 1979). We thank the Consortium on Chicago School Research for providing the Rasch measures used in this analysis.

they were part of bilingual or special education programs. Finally we exclude an additional 8,080 students who left the system or were placed in self-contained special education classes the following year because these students cannot be categorized as promoted or retained and generally do not have future test scores.

Table 1 presents summary statistics for the group of 147,894 students who experienced the accountability policy from 1997 to 1999. Chicago public school students are disproportionately minority and extremely low achieving compared with a national sample. Roughly 85 percent of Chicago students are black or Hispanic and the same fraction received free or reduced price lunches. Students who are retained are even more likely to be from minority backgrounds and low-income families.

Given the low achievement levels in the CPS, the promotional policy applied to a substantial proportion of students. Over 40 percent of third graders failed to meet the promotional standards from 1997 to 1999 as did about 30 percent of sixth graders. The reading exam proved to be a more difficult hurdle than the math exam, with nearly all students failing reading alone or both reading and math. Even after five to ten percent of students received waivers from the policy, 21 percent of third graders and 13 percent of sixth and eighth graders were required to repeat a grade.

## 5. Empirical Strategy

### *5.1. Identification*

In order to determine the impact of a remedial education program (or treatment) on student achievement, we would like to estimate the following relationship:

---

[7] Note that we include pre-policy cohorts in our sample. We will discuss later the reason for this inclusion.

(1) $$Y_{i,t+1} = \mathbf{B}\mathbf{X}_{i,t} + \beta(Treat)_{i,t} + u_i + \varepsilon_{i,t+1},$$

where Y is the outcome, X is a vector of demographic and past performance variables, and *Treat* is a binary variable that takes on a value of one if a student receives some type of treatment and zero otherwise. $u$ represents unobserved (to the researcher) student ability, $\varepsilon$ is an error term, and $t$ and $i$ are time and individual subscripts respectively. The primary concern is that because of selection on the part of students, teachers and parents, $\text{cov}(Treat, u) \neq 0$ which will bias OLS estimates of the treatment effect.

However, by tying promotional decisions to performance on standardized tests, the Chicago policy created a highly non-linear relationship between a student's current achievement and his or her probability of attending summer school or being retained. Figure 2 illustrates this relationship for third and sixth graders from 1997 to 1999. Roughly 90 percent of students who passed math but scored just below the cutoff in reading received some remedial treatment while almost no one who passed math and scored at or above the cutoff in reading attended summer school or was retained.

Assuming that unobservable characteristics do not vary discontinuously around the cutoff, the promotional decision rule provides exogenous variation in the treatment. Because treatment is perfectly correlated to observable characteristics, it is orthogonal to unobserved characteristics. One can thus identify the impact of these programs by simply comparing students who scored just below and just above the promotional cutoff. For example, if students who missed the cutoff (and were thus required to attend summer school) learned much more than students who just made the cutoff (and thus avoided summer school), then one might conclude that summer school had a positive impact on student achievement.

This strategy is often referred to as a regression discontinuity design. In one of the first papers to introduce this design, Thistlethwaite and Campbell (1960) utilized the fact that National Merit Awards are given on the basis of whether a test score exceeds a threshold to estimate the effect of the award on a student's other scholarship receipt and college aspirations. Others that have utilized this technique include Berk and Rauma (1983), Trochim (1984), Black (1999), Angrist and Lavy (1999), Hahn et. al. (1999), and Guryon (2000).

The fundamental assumption behind regression discontinuity techniques is that unobserved characteristics vary continuously (around the point of the cutoff) with the observable characteristic used to determine treatment. This assumption may not hold if individuals can influence their position relative to the cutoff. However, we believe that this type of intentional manipulation is implausible in our case. While a student may purposely miss many or all of the exam questions, it is unlikely that he or she would have the incentive or ability to marginally

change her score near the cutoff (e.g., intentionally scoring a 2.7 instead of a 2.8) because of the uncertainty regarding both performance and the grading metric.[10]

In the case of a sharp discontinuity, where performance exceeding a predetermined threshold perfectly predicts treatment, continuity of unobserved characteristics is sufficient to allow identification. In some cases, however, treatment may be partly determined by other factors, leading to a "fuzzy" discontinuity. For example, roughly 3 percent of students who scored below the cutoffs in June received waivers from summer school and about 14 percent of students in summer school received waivers in August. In addition, a small percentage of students who passed the exams were retained because of course failure or poor attendance. If waivers were distributed randomly, or on the basis of factors that are not correlated with future outcomes, then this would not present a problem. However, students who received waivers differ from their peers over several observable characteristics, raising a concern that these students differed along unobservable dimensions as well.

Even with waivers, as long as the probability of treatment changes discontinuously at the cutoff, it is possible to determine the treatment effect by comparing mean outcomes of individuals in a narrow range on either side of the cutoff. One merely needs to scale the difference in outcomes by the difference in the probability of treatment. If, however, the probability of treatment drops over a *range* around the cutoff, it may not be possible to identify the treatment effect by simply comparing individuals to the left and the right of the cutoff. We can, however, use a broader range of data to identify the effect. In essence, we can examine

[10] Teacher cheating is a more plausible candidate for such intentional manipulation. There is some evidence that cheating has increased as a result of the accountability policies instituted in 1996 and is more common in the promotional gate grades (Jacob and Levitt 2001). As a check on this, we estimated kernel densities of the test score distribution pre- and post-policy. If students could (and chose to) strategically influence their scores or teachers cheated in order to get their students above the cutoff, we would expect a discontinuity around the cutoff post-policy. We find no evidence of this.

whether performance drops (rises) in the range of performance where the probability of treatment is rapidly changing. To do so, we need to use data outside of this range to estimate the baseline relationship between initial performance and subsequent outcomes. In this case, we can use an instrumental variables strategy in which our instruments are non-linear terms of current test scores. These terms are highly correlated with the probability of treatment (as seen in Figure 1) but may not be directly correlated with future achievement. Because we use only the variation in treatment associated with observable performance, our point estimates should be unaffected by the correlation between treatment and unobserved characteristics.

One drawback of the IV approach described above is that it relies on knowing the functional form of the relationship between the outcome variable and the variable that determines treatment. In our case, this is the relationship between current test score and future performance. If, for example, the relationship is non-linear around the cutoff but we specify the function as linear, then the estimated treatment effect may simply pick up any underlying non-linearity in the achievement relationship. If the discontinuity is relatively sharp, then one can use a relatively narrow range of data so that a linear approximation is quite good. If the probability of treatment declines more slowly with observed test score, then we must rely more heavily on our functional form assumption.

We test the validity of this assumption in two ways. First, we examine the relationship between current test scores and future performance *prior* to the implementation of the accountability policy. Figure 3 shows that this relationship is indeed nearly linear, particularly in the range around the promotional cutoff. Second, we can test the robustness of our estimates by including second and third order polynomials of current test scores, in order to capture any

underlying non-linearity in the functional form. As we see in the next section, our estimates are robust to the inclusion of non-linear terms.

*5.2. Estimation*

A simple way to implement a regression discontinuity design is to compare the mean achievement gains of students just below the cutoff to those students at the cutoff. This can be represented mathematically with the following expression: $(\bar{Y}_{c,t+1} - \bar{Y}_{c,t}) - (\bar{Y}_{c-1,t+1} - \bar{Y}_{c-1,t})$, where $\bar{Y}$ represents mean achievement, the first subscript denotes performance relative to the reading cutoff ($c$ indicates students at the cutoff and $c$-1 denotes students just below the cutoff) and the second subscript denotes timing. Because every individual below the cutoff is not treated and some students above the cutoff are treated, we must scale the expression above by the difference in the probability of treatment associated with meeting the cutoff. Doing so yields the following difference-in-difference estimator:

$$\beta^{2sd} = \frac{(\bar{Y}_{c,t+1} - \bar{Y}_{c,t}) - (\bar{Y}_{c-1,t+1} - \bar{Y}_{c-1,t})}{\bar{T}_c - \bar{T}_{c-1}}, \tag{2}$$

where $\bar{T}$ is the mean probability of treatment. Notice that this difference-in-difference is equivalent to an IV estimate in which the only instrument is a dummy that takes on a value of one if the student passes the reading cutoff.

Though students below the cutoff and at the cutoff are similar, they are not identical. We might be worried that any difference in achievement gains between these groups reflects differences in initial ability rather than the influence of the treatment.[11] To address this concern, we can examine the performance of individuals just above the cutoff. If we assume that the

[11] In other words, high ability students may enjoy larger gains than students with lower initial ability.

typical difference in achievement between students at the cutoff and just below the cutoff is similar to the difference in achievement between those just above the cutoff and those at the cutoff, we can identify the treatment effect using the third difference estimator below:

$$\beta^{3d} = \frac{\left[\left(\bar{Y}_{c,t+1} - \bar{Y}_{c,t}\right) - \left(\bar{Y}_{c-1,t+1} - \bar{Y}_{c-1,t}\right)\right] - \left[\left(\bar{Y}_{c+1,t+1} - \bar{Y}_{c+1,t}\right) - \left(\bar{Y}_{c,t+1} - \bar{Y}_{c,t}\right)\right]}{\left(\bar{T}_c - \bar{T}_{c-1}\right) - \left(\bar{T}_{c+1} - \bar{T}_c\right)} \tag{3}$$

The third difference is equivalent to an IV estimate in which we control for a linear trend of reading ability and instrument for treatment using a dummy that takes on a value of one for students who exceed the cutoff.

*5.3. Implementing the Estimation Strategy in an IV Framework*

As we mentioned previously, the second and third difference approaches are simply IV strategies for estimating the treatment effect. We will implement the three-difference estimator in the following way. We will assume that our learning equation takes the following form:

$$Y_{i,t+1} = BX_{i,t} + \beta_1 rdge_{i,t} + \beta_2 (Treat)_{i,t} + u_i + \varepsilon_{i,t}, \tag{4}$$

where $Y$ is the academic outcome of interest, $X$ is a vector of demographic characteristics, *rdge* is the ITBS reading score, $u$ is unobserved ability, and $\varepsilon$ is an error term. The first stage is given by the following:

$$Treat_{i,t} = \Gamma X_{i,t} + \gamma_1 rdge_{i,t} + \gamma_2 1_{i,t}^{r} + \gamma_3 u_i + \eta_{i,t}, \tag{5}$$

where $1^{r}$ is a dummy variable that indicates that the reading score is above the cutoff, and $\eta$ is an error term. If we included no individual level covariates, this approach would be completely equivalent to the three-difference strategy summarized by equation (3). Implementing this approach using IV gives us the flexibility to include student level covariates as a check for the robustness of our estimates.

For students who passed math and failed reading in June, the probability of being retained does not change discontinuously at the cutoff as a function of August reading performance. Instead it drops off sharply in a range just below the cutoff that we will refer to as the marginal area. This can be observed in Figures 5 and 6. The fact that the probability of retention does not change discontinuously prevents us from using the same first stage relationship given by equation (5). Instead, while examining the retention treatment, we will estimate a first stage relationship of the following form:

$$(6) \qquad Treat_{i,t} = \Gamma X_{i,t} + \gamma_1 rdge_{i,t} + \gamma_2 1^m_{i,t} + \gamma_3 1^m_{it} rdge_{i,t} + \gamma_4 1^p_{i,t} + \gamma_5 1^p_{i,t} rdge_{i,t} + \gamma_6 u_i + \eta_{i,t},$$

where $1^m$ is a dummy variable that takes on a value of 1 if the student's reading score is in the marginal area and 0 otherwise. The other variables are as previously described. Away from the cutoff, increases in reading performance appear to have little effect on the probability of retention. This suggests that $\gamma_1$ and $\gamma_5$ are likely to be small. We interact the dummy variable for the marginal area with reading performance to take into account that for students just below the cutoff, small increases in performance lead to large reduction in the probability of being retained. If this is the case, $\gamma_3$ should be strongly negative. The learning equation is equivalent to that in equation (4).

## 6. The Net Effect of Summer School and Grade Retention on Student Achievement

Our analysis proceeds as follows. We first utilize the discontinuity created by the promotional cutoff associated with the *June* testing to estimate the *net* effect of summer school and grade retention. We next take advantage of the *August* cutoff to estimate the separate effect of grade retention, using the sample of students who were assigned to summer school solely on the basis of their June reading scores. For these students, the retention decision depended solely

on their August reading scores. Finally, we derive estimates of the separate effect of summer school, relying on the estimates of the net and retention effects described above.

The relatively sharp discontinuity between current achievement and the probability of attending summer school and possibly being retained permits one to visually identify the treatment effect. If these programs had a substantial net impact on subsequent academic achievement, we would expect to see a discontinuous or non-linear change in the average achievement level around the promotional policy cutoff. By plotting the probability of receiving some remedial treatment (summer school and retention) and future achievement against current test performance, Figures 4 and 5 allow this visual identification.[12] Each figure shows three relationships: (a) the probability of receiving remedial treatment (summer school or retention); (b) the reading performance of students in the following year; and (c) the math performance in the following year.

For example, in Figure 4, we see that the probability of attending summer school and being retained drops sharply at the cutoff for promotion in the third grade. At the same time, next year reading and math achievement drop sharply around the cutoff, suggesting that summer school and retention had a net positive effect for third graders. In contrast to third grade, the continuous linear trends in future performance among sixth graders (in Figure 5) suggest that the summer school and retention had no substantial effect on the performance of students near the cutoff.

Though the graphs lend transparency to the analysis, it is important to quantify the magnitude and the statistical precision of the estimates. We examine the subset of the student

[12] Note that these graphs reduce the dimensionality of the problem by limiting the sample to students who passed the standards in math. The thinness of data in some cells makes identification off of three-dimensional graphs quite difficult.

population for which only the reading cutoff is binding (i.e., students who passed math and only need to pass reading to avoid treatment).[13] Unless otherwise mentioned, the sample includes students who were in the third or sixth grades in 1997, 1998, or 1999. For all specifications, we include year fixed effects.

The results in Table 2 correspond to the effects displayed in Figures 4 and 5. The first column presents OLS results and the second column shows the IV results associated with the third difference specification described earlier. We see that there are no statistically significant differences between the OLS and IV estimates, although the IV point estimates tend to be somewhat larger than the OLS estimates. This suggests that June waivers were given randomly or on the basis of characteristics that were uncorrelated with future performance (to the extent that the IV estimates are in fact larger, one would conclude that waivers were given to students with positive unobservable characteristics). Column 3 shows IV estimates that control for a detailed set of student characteristics, including prior math and reading test scores, race gender, special education status, neighborhood poverty and free lunch status. The estimates in columns 2 and 3 are virtually identical, providing additional evidence that our instruments may be valid.

Summer school and grade retention have a positive net impact on third grade achievement in math as well as reading. In the first year, this effect was roughly 0.11-0.13 logits in the context of average third grade achievement gains of 0.68 and 0.42 logits in math and reading respectively. This means that summer school and grade retention increased student achievement roughly 20 percent of a year's worth of learning. By the second year after the

[13] We focus on students for whom the reading cutoff was binding because many more students failed on the account of reading than mathematics. In Table 3, we show results for students who passed reading and for whom the math cutoff was binding.

program, the effects had faded by roughly 25 to 40 percent, but were still statistically significant. This is consistent with the fadeout of program effects found in other evaluations (Barnett 1995). In sixth grade, the picture is much different. It appears that the net effect of summer school and grade retention for these older students was essentially zero in reading, and close to zero in mathematics, particularly by year two.

Table 3 examines the robustness of these estimates to specification and sample choices. The first robustness checks are designed to ensure that our findings are not sensitive to functional form assumptions. In the second row we control for third order polynomials in prior achievement. In the third row, we take advantage of data from pre-policy years to ensure that our findings are not driven by non-linearity in the relationship between initial performance and subsequent achievement. Intuitively, we subtract the pre-policy third difference estimate of the effect of surpassing the cutoff from the corresponding post-policy estimate. If the relationship between initial and subsequent performance is stable over time, this fourth difference will ensure that our findings are not driven by non-linearity.[14] The fourth row shows estimates using pre-policy data and controlling for polynomials of prior ability. While these estimates differ slightly from the baseline findings, the differences are not significant. Furthermore, the overall pattern of results stays the same.

The final three rows of Table 3 show net effect estimates when we modify the sample under examination. First, we include only those students with the same test forms to eliminate problems associated with initial performance measures that may not be comparable across

[14] We implement this strategy by performing IV in which we control for the dummy variable that indicates a student surpassed the cutoff and the interaction of this term with reading performance. Our instruments become the dummy variable and interaction term multiplied by another variable that indicates whether the cohort was exposed to the accountability policy.

students.[15] Next, we examine only those students with consistent grade patterns. This addresses problems of measurement error that may be attributable to the miscoding of grades. Neither of these two sample restrictions has any substantive effects on the estimated net effect. As a final robustness check, we examine those students who passed reading in June and were marginal in math.[16] The effect size is slightly larger for third grade reading and somewhat smaller for third and sixth grade math. Overall, these results are similar to our baseline estimates.

Table 4 examines the heterogeneity of two-year net effects across years and student subgroups. The first row shows the aggregate estimates taken from column 3 of Table 2. Examining the three cohorts of student to experience the policy, it appears that the 1997 cohort of third graders appears to have experienced somewhat larger positive effects than the later groups, although these differences are not statistically significant. Top achieving students appear to have the experienced the largest gains, although these estimates also have large standard errors. This may be due to the fact that these students attended summer school, but were not retained. There are no significant and consistent differences across race, gender or SES. Overall, the net treatment effect of summer school and grade retention appear fairly homogenous.

## 7. The Effect of Grade Retention

While the June discontinuity allows us to estimate the net treatment effect, the prior literature suggests that the effects of summer school and retention may be quite different. Fortunately, the structure of the accountability program in Chicago provides an opportunity to

[15] Different test forms are used for different years. The inclusion of year fixed effects addresses this concern.
[16] We can do this because of the two-dimensional nature of the cutoff. We do not emphasize these results because few students were treated on the basis of math—this reduces our sample size and the precision with which we can estimate the treatment effects.

separately identify the causal effect of grade retention. In order to advance to the next grade, the maximum of a student's June and August scores must exceed a predetermined cutoff in both reading and math. The discontinuity generated by the August cutoff allows us to estimate the impact of grade retention for the group of students who attended summer school.

Just as we did in estimating the net effects, we can reduce the dimensionality of the problem by considering only students who passed one subject in June and thus only had to pass the other subject in August. This is even more important in the case of the retention estimates because it allows us to focus on a student's August score alone. If we instead focus on the maximum of June and August test scores, it is likely that students who scored just above the cutoff will differ from students who scored just below the cutoff, thus violating the central assumption of a regression discontinuity design. This is true because students whose June scores exceeded the cutoff never attended summer school, which means that summer school students whose maximum (June and August) score exceeded the cutoff must have improved over the summer. It is likely that these students were more motivated than those students whose maximum score did not exceed the cutoff. For this reason, the analysis below focuses on the subset of students who passed math and failed reading in June.[17]

Figures 6 and 7 show the probability of retention and future academic outcomes as a function of August reading scores. While the probability of retention does not drop sharply at the exact point of the cutoff, we can see that it rapidly decreases over a narrow range of values just below the cutoff. We will refer to this range as the marginal area. Figure 6 shows that, for third grade students, future performance is flat or decreasing in the marginal area, consistent with

[17] We could also examine individuals who passed reading and failed math in June. In practice, however, very few of these students end up near the cutoff in August.

a positive effect of grade retention. In contrast, Figure 7 presents little evidence that retention has any benefit for sixth grade students.

Because of the fuzzy discontinuity, we cannot limit our analysis to students immediately above and below the cutoff. However, if we use too broad a range of data, students at the extreme ends of the distribution are unlikely to be comparable, forcing us to rely heavily on our covariates to control for the differences between students. As a compromise, we focus on a subset of children who scored relatively close to the cutoff on the August exam, including children who scored from 1 GE below the cutoff to 0.5 GE above the cutoff.[18] To ascertain whether students are comparable (conditional on August reading performance), we first estimate the retention treatment effect without using additional covariates. We then estimate the effect controlling for a rich set of prior achievement measures and demographic characteristics. If the results are insensitive to the inclusion of these covariates, it is likely that the students to the left and right of the cutoff are comparable. In addition, we test whether our estimates are robust to changes in the range of students in the sample.

Table 5 reports the coefficients from the first stage, which are as expected. In particular, the coefficient of the marginal reading interaction term is strongly negative for both third and sixth grades. This confirms that small changes in reading performance in the marginal area are associated with large reductions in the probability of being retained. The instruments also have strong predictive power; the F-statistic of the instruments is 236 for the third grade cohorts and 150 for the sixth grade cohorts.

Table 6 contains the estimated retention treatment effects for students in the third and sixth grades respectively. We see many of the same patterns as we did for the net effects. The

IV point estimates (shown in column 2) are somewhat larger than the OLS estimates (shown in column 1), suggesting that waivers were given to students with positive unobservable characteristics (note that these differences are not statistically significant.) The third column shows IV estimates that control for student characteristics. These estimates are similar to those in column 2, suggesting that the students in our analysis are comparable once we control for August reading ability.

The point estimates in Table 6 suggest that retention may not have as powerful a negative effect on academic achievement as commonly cited in the literature. The IV estimates indicate that being retained in the third grade actually *increases* performance the following year by 0.17 logits in reading and 0.23 logits in math. These treatment effects correspond to increases in achievement of 41 and 33 percent of the average annual gain. By the second year following retention, the math effect has decreased substantially, but is still significant. The two-year reading effect, however, is not statistically different than zero.

While it appears that the retention effects have become more negative by the second year, changes in the policy effect over time are confounded by the changes in student incentives from grade to grade. One year after third grade, retained students (i.e., those who are repeating the third grade) face high stakes testing again while promoted students (now in the fourth grade) do not. Two years later, the majority of retained students as well as promoted students face little incentive to perform well on the ITBS exams. For this reason, the two-year estimates provide the most accurate view of the retention effects for third graders.

When we compare sixth grade students one year later, we find no statistically significant differences between the performance of retained and promoted students, despite the fact retained

---

[*] We include a broader range of data below the cutoff in order to effectively estimate the effect of reading

students faced high-stakes testing and the promoted students did not. It may be the case that negative retention effects were offset by the positive incentive effects. When we compare these students after two years, the incentives are reversed. Students who had been promoted in sixth grade are most likely in eighth grade, facing a high-stakes exam once again, whereas retained students are most likely in seventh grade, facing a low-stakes exam. For this reason, the two-year effects probably reflect an upper bound on any negative retention effects. And, in fact, we do find that retained students score roughly 0.15 logits (27 percent of an annual learning gain) lower than promoted students. However, there is no significant difference between the math achievement of retained and promoted students.

Table 7 shows the sensitivity of our results to functional form assumptions and choice of samples. Rows 2 and 3 show how our estimates vary as we control for second and third order polynomials of initial performance. Our estimates appear robust to these changes.[19] Rows 4, 5, and 6 show how our estimates change as we use data from broader and narrower ranges around the cutoff. The results do vary somewhat depending on our sample but the differences are not significant and do not change the pattern of results. The final row shows estimates of the retention treatment effect when we examine students who passed reading and failed math in June. Our findings in some cases are quite different when we examine this sample. The difference is significant only for third grade math, however. In all cases, the standard errors of these estimates are quite large making it difficult to draw strong conclusions.

Table 8 reports the two-year retention effects for various cohorts and sub-samples. Several interesting patterns are evident. First, it appears that the effect of grade retention has

---

performance on the probability of retention at and below the marginal area.

[19] Because the ITBS exam was not given in August prior to 1997, there it is not possible to observe the counterfactual relationship between August scores and subsequent performance.

improved for later cohorts, although relatively large standard errors make it impossible to state this with a high level of confidence. This pattern, however, is consistent with the introduction of an after-school tutoring program for retained students starting in 1998. There are no consistent or statistically significant differences in the magnitude of the treatment effects by race, gender, SES, or prior achievement.

**8. The Effect of Summer School**

In Section 6 we estimated the net effect ( $\beta_N$ ) of summer school and grade retention. Using information on the magnitude of the retention treatment effects, we can now back out an estimate of the summer school treatment effect. Assuming homogeneous treatment, the net effect can be represented in the following way:

(7) $\beta_N = \beta_S + \beta_R P_R$,

where $\beta_S$ and $\beta_R$ are the summer school and retention treatment effects respectively and $P_R$ is the probability of being retained conditional on attending summer school. Hence, to determine the separate effect of summer school, we must obtain estimates of the net effect of summer school and grade retention ( $\beta_N$ ), the probability of retention conditional on attending summer school ( $P_R$ ), and the separate effect of grade retention ( $\beta_R$ ).

For the net effects, we will use the third difference estimates presented in Table 2. While the probability of retention is simple to calculate empirically, it is important to use the same population that was used to estimate the net effect. Recall that this group consisted of students who in June passed the promotional cutoff in math and scored just below or just above the promotional cutoff in reading (which we therefore refer to as the June sample). It can be shown

that the conditional probability of retention consistent with our third difference estimate of the net effect is the following:

$$(8) \qquad \hat{P}_R = \frac{\left(\overline{R}_c - \overline{R}_{c-1}\right) - \left(\overline{R}_{c+1} - \overline{R}_c\right)}{\left(\overline{T}_c - \overline{T}_{c-1}\right) - \left(\overline{T}_{c+1} - \overline{T}_c\right)},$$

where $\overline{R}$ is the fraction retained and the other variables are as previously described. This estimate is the probability of being retained for those students who went to summer school because they missed the cutoff.

Finally, we must obtain an estimate of the retention effect, once again taking care to use the same population that was used to determine the net effect. Unfortunately, the sample used to estimate the retention effects in Section 7 is somewhat different than the June sample. The retention effects were estimated using the sample of students who passed the math, but not the reading, cutoff in June, and then scored within a limited range around the reading cutoff (both above and below) in August (which we refer to as the August sample). In general, this is a lower ability group than the population used to estimate the net effects. To the extent that the effect of retention is different across the two groups, our estimates of the summer school effect will be biased. While the nature of our identification strategy prevents us from using the same sample to estimate the net and retention effects, we can attempt to place reasonable bounds on the summer school effect.

The middle column in Table 9 presents the summer school effect under the assumption that the retention effect is identical for the June and August samples. Using the retention effects presented in Table 6, the implied two-year summer school treatment effects range from 0.03 to 0.07 logits. For third grade students, summer school increases reading and math achievement

two year later by roughly 12 percent of the average annual learning gain. For sixth graders, the effects are roughly half as large.

As we mentioned previously, it may be that the retention treatment effects are different for students near the cutoff in August than for students near the cutoff in June. To place reasonable bounds on the potential summer school effects, we subtract 0.25 logits from the retention treatment effect for the upper estimates and add 0.25 logits for the lower estimates. Depending on our assumption regarding the magnitude of the retention treatment effect, the implied summer school effects can vary substantially. Despite this, it appears that even under very pessimistic assumptions, summer school improves performance in mathematics. In reading, the lower bound estimates are close to zero. Conversely, the upper bound estimates are substantial for reading and for mathematics.

## 9. Conclusions

As school districts impose tougher standards on students and increasingly hold schools accountable for their performance, there will be a growing need to find effective remedial education programs to help low-achieving students. The evidence presented in this paper suggests that summer school and grade retention have a modest but positive net impact on achievement scores for third grade students, but little if any effect on academic achievement for sixth grade students. These findings are robust to various specifications, and do not appear to vary substantially across different groups. However, these programs do appear to have a larger effect on math than reading achievement.

For third grade students, the net effect is a combination of benefits from both summer school and grade retention. Contrary to conventional wisdom and prior research, we find that

retention may actually increase academic achievement for low-achieving third graders. For sixth grade students, the zero net effects appear to mask a small positive summer school effect and a negative retention effect. However, it is also possible that the negative effects for sixth graders are driven by the higher incentives that promoted students face two years later.

These finding suggest that school districts that implement policies aimed at increasing standards and accountability can utilize summer school and (to a lesser extent) grade retention to support students who cannot initially meet the new standards. One limitation of this analysis is that we were only able to follow students for several years following summer school and/or retention. It is possible that the summer school benefits will fade and that the long term effects of retention will be worse than the short term effects. As more time passes from the implementation of the program, it will be possible to examine these possibilities in greater detail.

**References**


Alexander, K. L., D. R. Entwisle, et al. (1994). On the Success of Failure: A Reassessment of the Effects of Retetenion in the Primary Grades. New York, University of Cambridge Press.

Angrist, J. D. and V. Lavy (1999). "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement." Quarterly Journal of Economics **114**(2): 535-75.

Barnett, W. S. (1995). "Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes." The Future of Children 5(3): 25-51.

Berk, R. A. and D. Rauma (1983). "Capitalizing on Nonrandom Assignment to Treatments: A Regression-Discontinuity Evaluation of a Crime-Control Program." Journal of the American Statistical Association **78**(381): 21-28.

Black, S. (1999) "Do Better Schools Matter? Parental Valuation of Elementary Education." Quarterly Journal of Economics **114**: 577-599.

Cooper, H., K. Charlton, et al. (2000). Making the Most of Summer School: A Meta-Analysis and Narrative Review. Monograph series of the Society for Research in Child Development. Malden, MA.

Dworkin, A. G., J. Lorence, et al. (1999). Elementary School Retention and Social Promotion in Texas: An Assessment of Students who Failed the Reading Section of the TAAS. Houston, University of Houston.

ECS (2000). ECS State Notes. Education Commission of the States (www.ecs.org).

Eide, E. R. and M. H. Showalter (*forthcoming*). "The Effect of Grade Retention on Educational and Labor Market Outcomes." Economics of Education Review.

Fine, M. (1991). Framing Dropouts: Notes on the Politics of an Urban Public High School. Albany, N.Y., SUNY Press.

Grissom, J. B. and L. A. Shepard (1989). Repeating and Dropping Out of School. Flunking Grades: Research and Policies on Retention. L. A. Shepard and M. L. Smith. New York, The Falmer Press: 34-63.

Guryan, J. (2000) "Does Money Matter? Regression Discontinuity Estimates from Education Finance Reform in Massachusetts." University of Chicago Working Paper.

Hahn, J., P. Todd, et al. (1999). Evaluating the Effect of an Anti-Discrimination Law Using a Regression-Discontinuity Design. NBER Working Paper #7131. Cambridge, MA.

Hanushek, E. A. (1996). "School Resources and Student Performance." In Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success. ed. Burtless, G. Washington D.C.: Brookings Institution Press.

Hedges, L. V. and R. Greenwald (1996). "Have Times Changed? The Relation between School Resources and Student Performance." In Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success. ed. Burtless, G. Washington D.C.: Brookings Institution Press.

Holmes, C. T. (1989). Grade Level Retention Effects: A Meta-Analysis of Research Studies. Flunking Grades: Research and Policies on Retention. L. A. Shepard and M. L. Smith. New York, The Falmer Press: 16-33.

Hoxby, C. M. (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." Quarterly Journal of Economics **115** (4): 1239-1285.

Jacob, B. A. and S. D. Levitt (2001). High-Stakes Testing and Teacher Cheating in the Chicago Public Schools. University of Chicago, Working Paper.

Jacob, B.A. (2001). The Impact of High-Stakes Testing on Student Achievement: Evidence from Chicago. Working paper, John F. Kennedy School of Government, Harvard University.

Karweit, N. L. (1991). Repeating a Grade: Time to Grow or Denial of Opportunity. Baltimore, Maryland, Center for Research on Effective Schooling for Disadvantaged Students.

Katz, L. F. and K. M. Murphy (1992). "Changes in Relative Wages 1963-1987: Supply and Demand Factors." Quarterly Journal of Economics **107** (1): 35-78.

Krueger, A. B. (1999). "Experimental Estimates of Education Production Functions." Quarterly Journal of Economics **114**:497-532.

Murnane, R. J., J. B. Willet, and F. Levy (1995). "The Growing Importance of Cognitive Skills in Wage Determination." The Review of Economics and Statistics **77** (2):251-266.

Pierson, L. H. and J. P. Connell (1992). "Effect of Grade Retention on Self-System Processes, School Engagement and Academic Performance." Journal of Educational Psychology **84**: 300-307.

Pipho, C. (1999). "Summer School: Rx for Low Performance." Phi Delta Kappan **81**(September): 7.

Roderick, M. (1994). "Grade Retention and School Dropout: Investigating the Association." American Educational Research Journal **31**(4): 729-759.

Roderick, M., J. Nagaoka, et al. (2000). Update: Ending Social Promotion. Chicago, IL, Consortium on Chicago School Research.

Rumberger, R. W. (1987). "High School Dropouts: A Review of Issues and Evidence." Review of Educational Research **57**: 101-121.

Schulz, E. M., R. Toles, et al. (1986). Association of Dropout Rates with Student Attributes. American Educational Research Association, San Francisco, CA.

Shepard, L. A. and M. L. Smith (1989). Flunking Grades: Research and Policies on Retention. New York, N.Y., The Falmer Press.

Thistlewaite, D. and D. Campbell (1960). "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment." Journal of Educational Psychology **51**: 309-317.

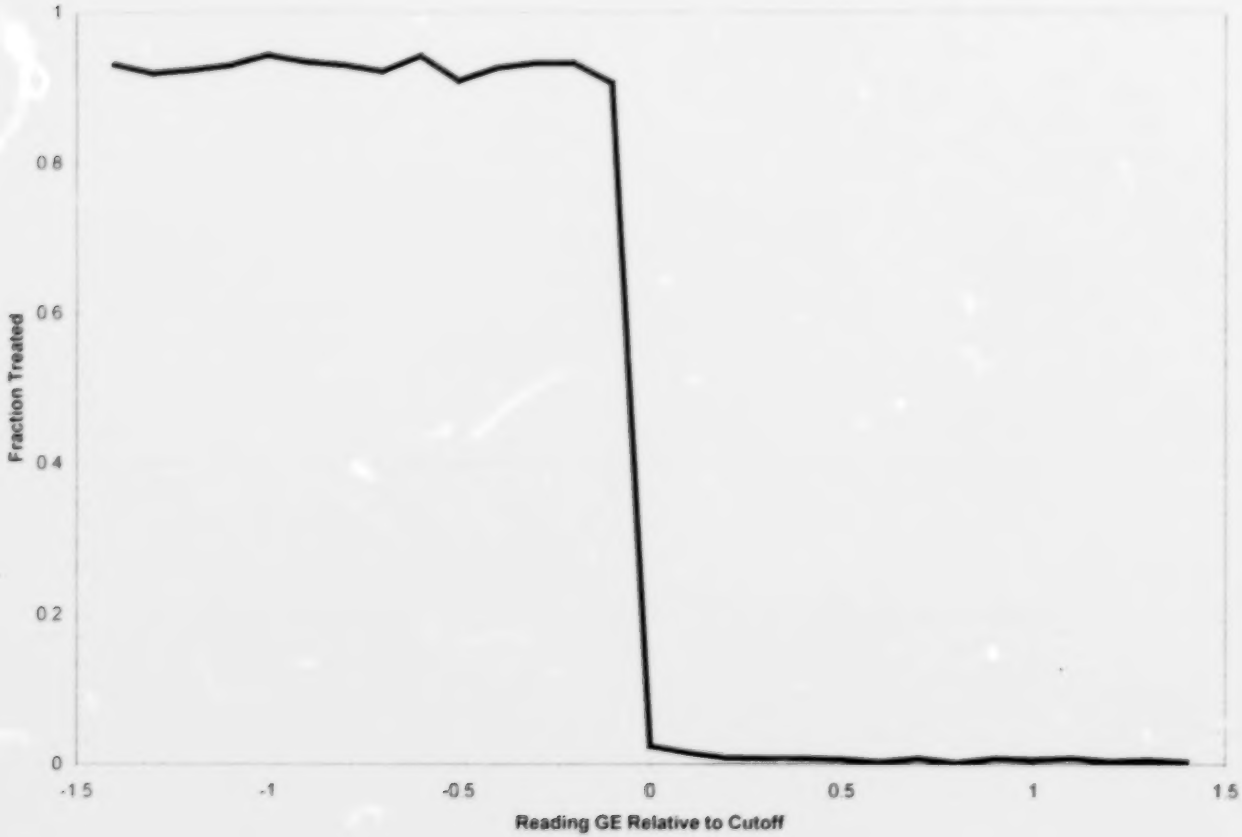
Topel, R. (1999). "Labor Markets and Economic Growth." in Handbook of Labor Economics. ed. Ashenfelter, O. and Card, D. Amsterdam: Elsevier Science B.V. pp. 2943-2984.

Trochim, W. (1984). Research Design for Program Evaluation: The Regression-Discontinuity Approach. Beverley Hills, CA, Sage Publications.

Wright, B. and M. H. Stone (1979). Best Test Design. Chicago, IL, MESA Press.

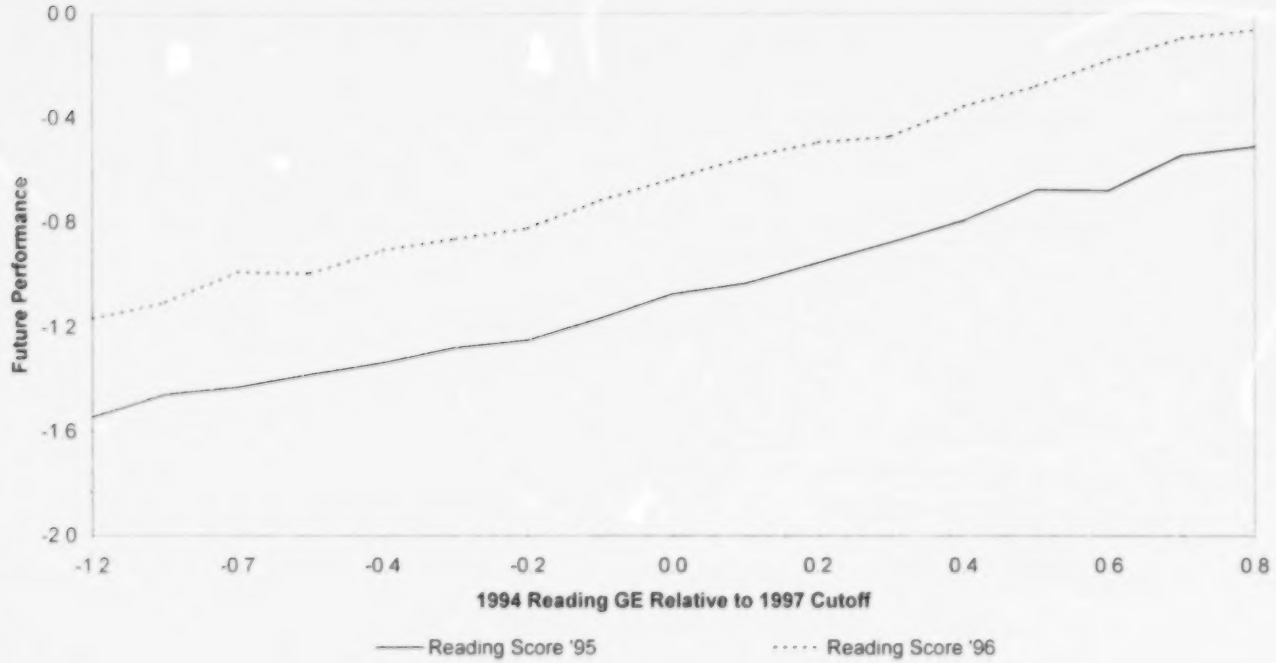
**Figure 1: Student Progress Under the Chicago Accountability Policy**

**Figure 2: The Relationship Between June Reading Scores and the Probability of Attending Summer School or Being Retained**



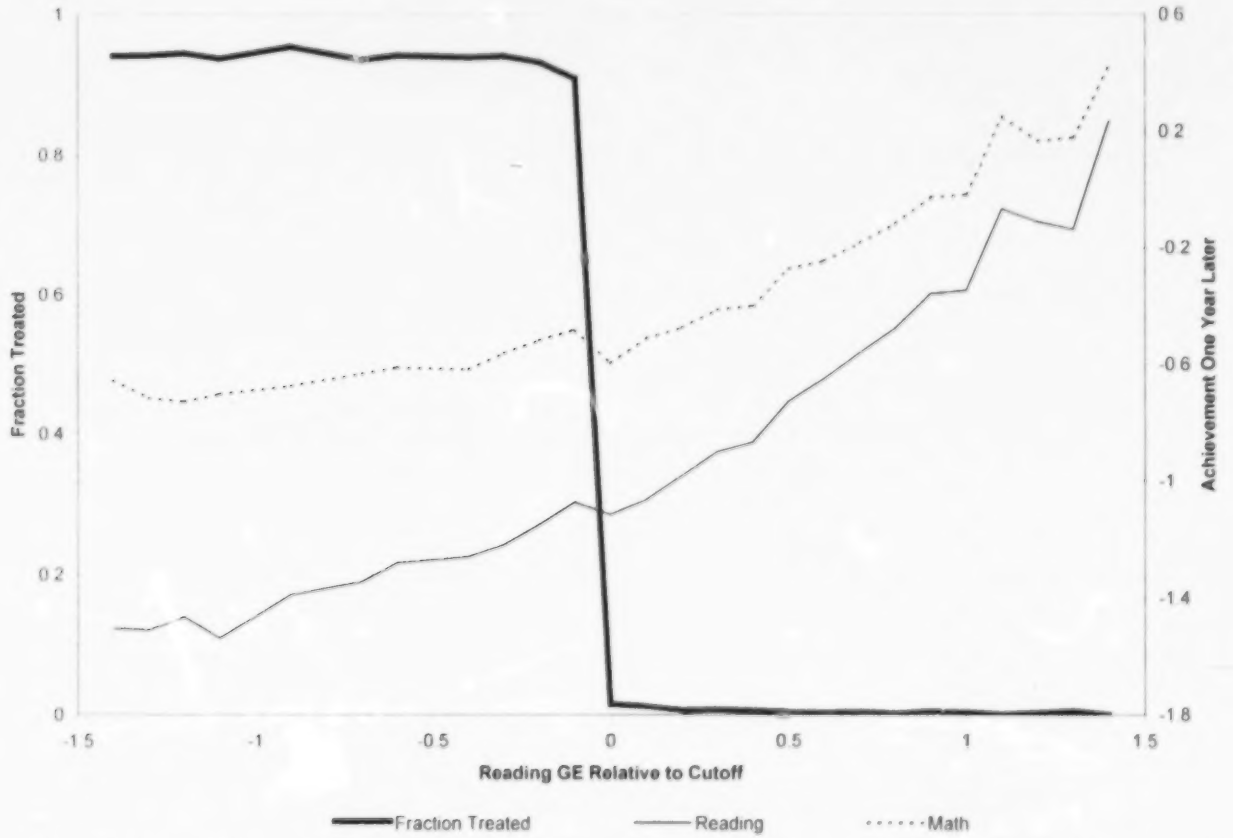
Notes: Sample of third and sixth grade students from 1997 to 1999 whose June math score exceeded the promotional cutoff but whose June reading score did not.

**Figure 3: The Relationship Between Current and Future Reading Performance of Third Grade Student Prior to the Accountability Policy**



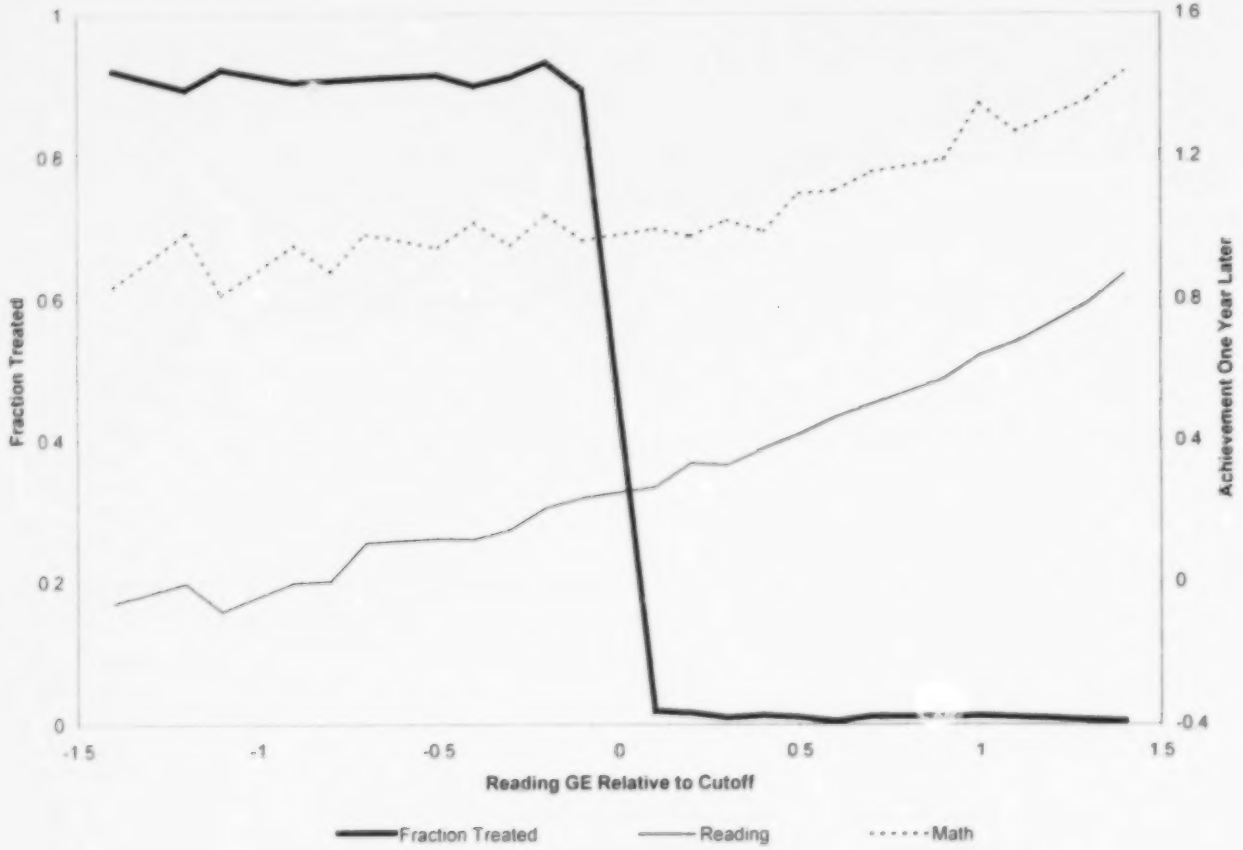

Notes: Sample includes third grade students in 1994 whose June math score exceeded what was the promotional cutoff in 1997.

**Figure 4: The Relationship between Reading and Math Performance and June Reading Performance for Third Grade Students.**



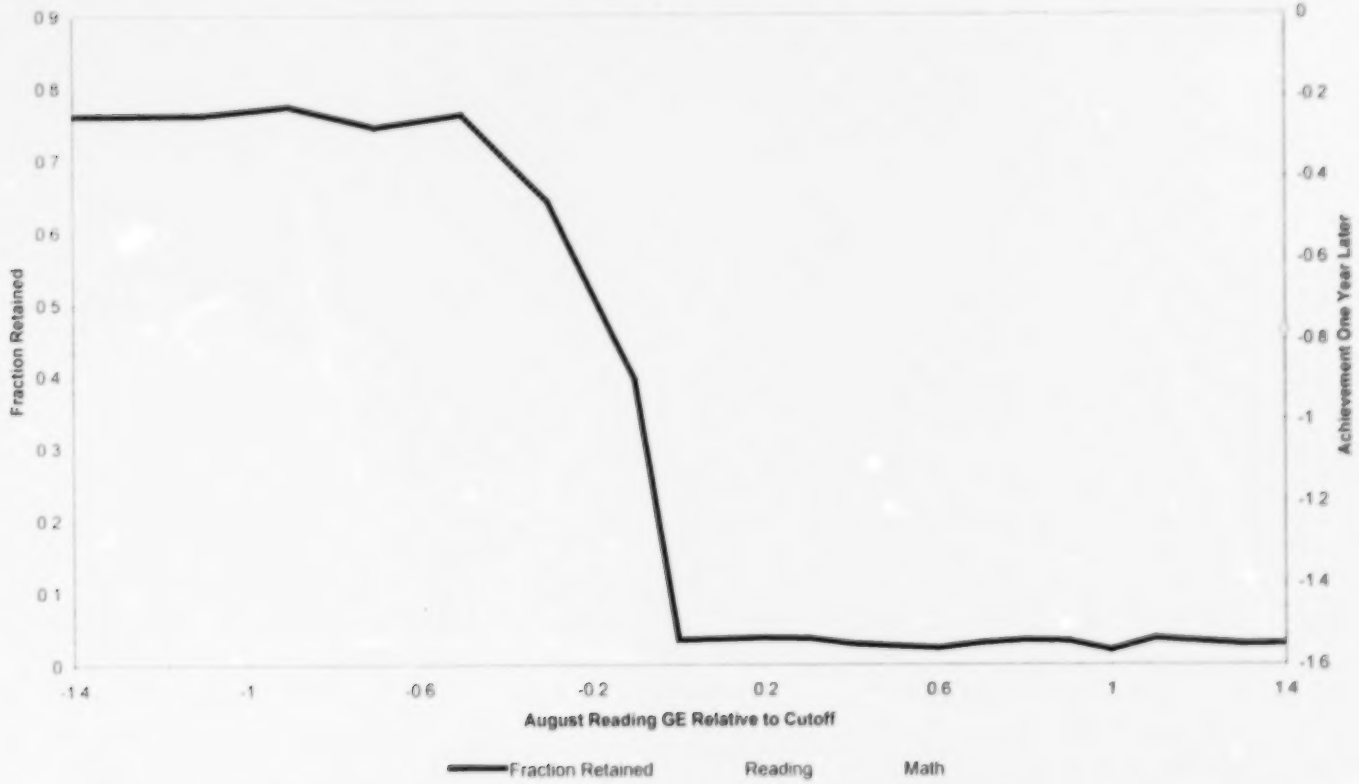

Notes: Sample of third grade students from 1997 to 1999 whose June math score exceeded the promotional cutoff but whose June reading score did not.

**Figure 4: The Relationship between Reading and Math Performance and June Reading Performance for Sixth Grade Students.**



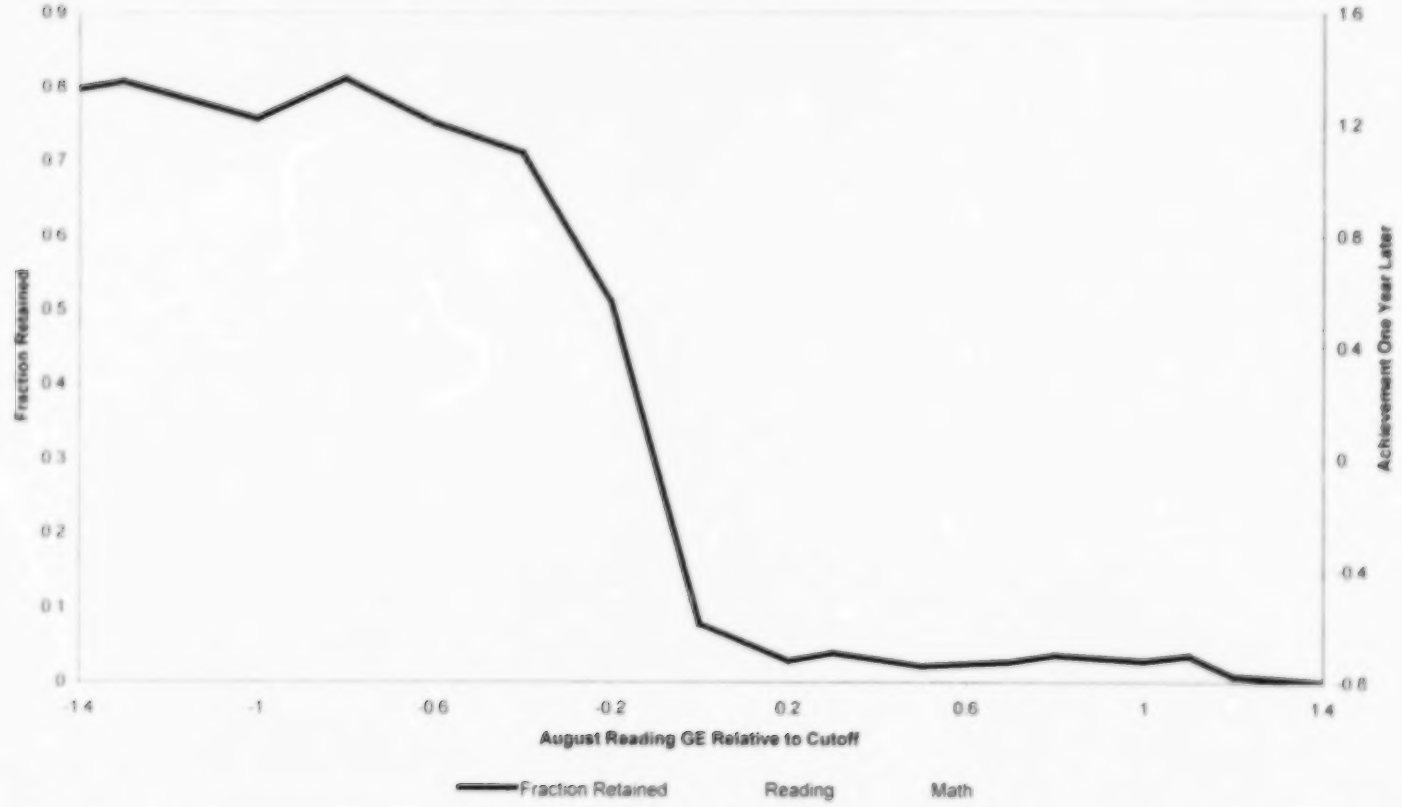

Notes: Sample of third grade students from 1997 to 1999 whose June reading score exceeded the promotional cutoff but whose June math score did not.

**Figure 6: Relationship Between August Reading and Next Year Reading and Math Performance for Third Grade Students**




Notes: Sample of third grade students from 1997 to 1999 whose June reading score exceeded the promotional cutoff but whose June reading score did not.

**Figure 7: Relationship Between August Reading and Next Year Reading and Math Performance for Sixth Grade Students**




Notes: Sample of sixth grade students from 1987 to 1989 whose June math score exceeded the promotional cutoff but whose June reading score did not.

**Table 1**
**Summary Statistics**

| | Third Grade | | | Sixth Grade | | |
|---|---|---|---|---|---|---|
| | Total | Failed Promotion Cutoff | Retained | Total | Failed Promotion Cutoff | Retained |
| ***Student Characteristics*** | | | | | | |
| Black | 0.713 | 0.822 | 0.844 | 0.553 | 0.645 | 0.678 |
| Hispanic | 0.174 | 0.136 | 0.123 | 0.318 | 0.314 | 0.288 |
| Male | 0.489 | 0.528 | 0.551 | 0.480 | 0.518 | 0.535 |
| Black Male | 0.347 | 0.435 | 0.467 | 0.261 | 0.334 | 0.364 |
| Hispanic Male | 0.087 | 0.070 | 0.069 | 0.156 | 0.162 | 0.152 |
| Age | 9.379 | 9.438 | 9.406 | 12.353 | 12.471 | 12.448 |
| Free Lunch | 0.805 | 0.904 | 0.927 | 0.774 | 0.890 | 0.905 |
| Reduced Price Lunch | 0.075 | 0.043 | 0.031 | 0.088 | 0.052 | 0.041 |
| Currently in Bilingual Program | 0.023 | 0.021 | 0.020 | 0.112 | 0.206 | 0.197 |
| Formerly in Bilingual Program | 0.135 | 0.087 | 0.076 | 0.228 | 0.111 | 0.090 |
| Special Education | 0.033 | 0.038 | 0.039 | 0.019 | 0.023 | 0.025 |
| Living with Relatives | 0.113 | 0.084 | 0.076 | 0.121 | 0.105 | 0.100 |
| Living in Foster Care | 0.067 | 0.086 | 0.088 | 0.045 | 0.062 | 0.067 |
| ***Experience Under the Accountability Policy*** | | | | | | |
| Passed in June | 0.573 | 0.000 | 0.012 | 0.690 | 0.000 | 0.035 |
| Failed Math Only | 0.044 | 0.102 | 0.041 | 0.060 | 0.193 | 0.135 |
| Fail Reading Only | 0.199 | 0.465 | 0.366 | 0.147 | 0.473 | 0.352 |
| Failed Math and Reading | 0.185 | 0.432 | 0.581 | 0.104 | 0.334 | 0.478 |
| June Waiver | 0.028 | 0.065 | 0.000 | 0.028 | 0.089 | 0.000 |
| Assigned to Summer School | 0.401 | 0.933 | 1.000 | 0.286 | 0.908 | 1.000 |
| August Waiver | 0.058 | 0.135 | 0.000 | 0.038 | 0.121 | 0.000 |
| Promoted | 0.791 | 0.517 | 0.000 | 0.873 | 0.606 | 0.000 |
| Retained | 0.209 | 0.483 | 1.000 | 0.127 | 0.394 | 1.000 |
| ***School Performance*** | | | | | | |
| Base year math score | -1.121 | -1.916 | -2.114 | 0.772 | -0.045 | -0.210 |
| | (1.066) | (0.726) | (0.695) | (0.917) | (0.589) | (0.563) |
| Year one math score | -0.440 | -1.039 | -1.210 | 1.264 | 0.566 | 0.348 |
| | (0.984) | (0.804) | (0.843) | (0.899) | (0.641) | (0.665) |
| Year two math score | 0.118 | -0.491 | -0.668 | 1.739 | 1.074 | 0.785 |
| | (0.982) | (0.760) | (0.761) | (0.856) | (0.661) | (0.676) |
| Base year reading score | -1.205 | -2.111 | -2.280 | 0.162 | -0.755 | -0.836 |
| | (1.060) | (0.560) | (0.540) | (0.951) | (0.466) | (0.479) |
| Year one reading score | -0.796 | -1.440 | -1.610 | 0.727 | -0.059 | -0.334 |
| | (1.010) | (0.721) | (0.752) | (0.970) | (0.638) | (0.640) |
| Year two reading score | -0.350 | -0.997 | -1.212 | 1.291 | 0.540 | 0.213 |
| | (1.011) | (0.704) | (0.688) | (0.962) | (0.695) | (0.700) |
| Number of Observations | 74,260 | 31,738 | 15,514 | 73,634 | 22,861 | 9,332 |

Notes to Table 1: The sample contains all students who were enrolled in the appropriate grades between the Spring of 1997 and the Spring of 1999.

**Table 2**
**The Net Effect of Summer School and Grade Retention on Student Achievement**

| Dependent Variables | Specification | | |
|---|---|---|---|
| | OLS | IV | IV |
| | (1) | (2) | (3) |
| **Third Grade** | | | |
| Reading | | | |
| One year (n=13,687) | 0.082 | 0.112 | 0.104 |
| | (0.019) | (0.026) | (0.025) |
| Two years (n=12,806) | 0.032 | 0.064 | 0.062 |
| | (0.020) | (0.027) | (0.026) |
| Math | | | |
| One year (n=13,664) | 0.155 | 0.132 | 0.136 |
| | (0.019) | (0.026) | (0.024) |
| Two years (n=12,802) | 0.066 | 0.087 | 0.095 |
| | (0.021) | (0.027) | (0.026) |
| **Sixth Grade** | | | |
| Reading | | | |
| One year (n=7,920) | -0.013 | 0.012 | 0.024 |
| | (0.022) | (0.029) | (0.027) |
| Two years (n=7,262) | -0.027 | -0.015 | 0.000 |
| | (0.024) | (0.032) | (0.030) |
| Math | | | |
| One year (n=7,904) | 0.056 | 0.077 | 0.077 |
| | (0.016) | (0.021) | (0.021) |
| Two years (n=7,249) | 0.007 | 0.018 | 0.019 |
| | (0.019) | (0.025) | (0.023) |
| Additional Performance and Demographic Covariates | No | No | Yes |

Notes for Table 2: The sample third and sixth grade students from 1997 to 1999 whose June math score was above the cutoff and whose June reading score was within + or - 0.2 grade equivalents of the cutoff. Year fixed effects are included in each model. The additional performance and demographic covariates include: one and two year prior achievement scores in math and reading (with missing values set to zero), along with variables that indicate whether these scores were missing, age, male, black, Hispanic, black*male, Hispanic*male, free lunch, reduced price lunch, special education participation, current bilingual participation, past bilingual participation, lives in foster care, lives with a non-parent relative, census block level social status and census block level poverty.

**Table 3**
**The Robustness of Two-Year Net Summer School and Retention Estimates to Sample and Specification Choice**

| Specification | Third Grade | | Sixth Grade | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| Baseline | 0.062<br>(0.026) | 0.09<br>(0.026) | 0.000<br>(0.030) | 0.019<br>(0.023) |
| Including polynomials in prior achievement | 0.120<br>(0.038) | 0.089<br>(0.037) | -0.021<br>(0.035) | -0.007<br>(0.028) |
| Including pre-policy cohorts (fourth difference estimates) | 0.029<br>(0.018) | 0.060<br>(0.017) | 0.003<br>(0.021) | 0.042<br>(0.017) |
| Including pre-policy cohorts and polynomials in prior achievement | 0.018<br>(0.018) | 0.062<br>(0.017) | -0.006<br>(0.024) | 0.025<br>(0.019) |
| Including Only Students with Common Test Form and Level | 0.060<br>(0.026) | 0.091<br>(0.025) | 0.003<br>(0.030) | 0.021<br>(0.023) |
| Including Only Students with Consistent Grade Patterns | 0.064<br>(0.026) | 0.094<br>(0.026) | -0.004<br>(0.030) | 0.018<br>(0.023) |
| Passed Reading and Marginal in Math | -0.007<br>(0.041) | 0.062<br>(0.033) | 0.058<br>(0.043) | 0.007<br>(0.033) |

Notes for Table 3: Each cell contains an estimate from a separate 2SLS regression that controls for all of the past performance and demographic characteristics listed in the notes to Table 2. For estimates in the last row, we use children who passed reading and were near the cutoff in math. For the other specifications we use children who passed math and were near the cutoff in reading.

**Table 4**
**The Heterogeneity of Two-Year Net Summer School and Probation Effects**

| | Third Grade | | Sixth Grade | |
|---|---|---|---|---|
| | **Reading** | **Math** | **Reading** | **Math** |
| Baseline estimates | 0.062 | 0.095 | 0.000 | 0.019 |
| | (0.026) | (0.026) | (0.030) | (0.023) |
| *Year* | | | | |
| 1997 Cohort | 0.093 | 0.154 | -0.001 | 0.022 |
| | (0.052) | (0.050) | (0.047) | (0.039) |
| 1998 Cohort | 0.032 | 0.065 | 0.007 | 0.025 |
| | (0.039) | (0.038) | (0.054) | (0.042) |
| 1999 Cohort | 0.058 | 0.085 | -0.006 | 0.020 |
| | (0.047) | (0.047) | (0.054) | (0.041) |
| *Prior Achievement*[a] | | | | |
| Bottom Quartile | 0.074 | 0.058 | -0.003 | -0.003 |
| | (0.056) | (0.055) | (0.057) | (0.045) |
| 2nd Quartile | 0.073 | 0.128 | -0.024 | 0.011 |
| | (0.045) | (0.043) | (0.044) | (0.034) |
| 3rd Quartile | 0.011 | 0.040 | 0.031 | 0.031 |
| | (0.048) | (0.048) | (0.060) | (0.048) |
| Top Quartile | 0.202 | 0.189 | -0.006 | 0.127 |
| | (0.095) | (0.090) | (0.216) | (0.175) |
| *Race, Gender & SES* | | | | |
| Black | 0.071 | 0.090 | -0.047 | -0.015 |
| | (0.030) | (0.025) | (0.040) | (0.031) |
| Hispanic | 0.053 | 0.117 | 0.049 | 0.061 |
| | (0.062) | (0.062) | (0.048) | (0.039) |
| White/Other | -0.018 | 0.082 | 0.175 | 0.092 |
| | (0.109) | (0.104) | (0.124) | (0.092) |
| Male | 0.034 | 0.077 | 0.002 | -0.011 |
| | (0.037) | (0.037) | (0.043) | (0.034) |
| Female | 0.086 | 0.112 | -0.008 | 0.047 |
| | (0.036) | (0.035) | (0.041) | (0.032) |
| Free Lunch | 0.061 | 0.094 | -0.005 | 0.018 |
| | (0.027) | (0.026) | (0.031) | (0.024) |
| No Free Lunch | 0.070 | 0.126 | 0.089 | 0.051 |
| | (0.094) | (0.095) | (0.119) | (0.091) |

Notes for Table 4: Each cell includes an estimate from a separate 2SLS regression (equivalent to the third difference estimates shown in Table 2 and described in the text) that controls for year fixed effects and all of the additional performance and demographic variables listed in the notes to Table 2. [a]Prior achievement is measured as the average math and reading score in second or fifth grade. Quartiles are determined on the basis of students in this sample. A small number of students who were missing second or fifth grade test scores are excluded from this categorization.

**Table 5**
**The Effect of August Test Performance on the Probability of Grade Retention**

| Independent variables | Third Grade | Sixth Grade |
|---|---|---|
| Math GE | -0.027**<br>(.008) | -0.013*<br>(0.009) |
| Reading GE | -0.024<br>(0.047) | 0.047<br>(0.077) |
| Marginal Reading | 2.451**<br>(0.178) | 3.387**<br>(0.472) |
| Marginal Reading*Reading GE | -1.031**<br>(0.074) | -0.733**<br>(0.102) |
| Passed Reading | -0.862**<br>(0.162) | -0.250<br>(0.410) |
| Passed Reading*Reading GE | 0.058<br>(0.063) | -0.098<br>(0.088) |
| Number of Observations | 7,623 | 4,552 |
| R-Squared | 0.477 | 0.504 |
| F-Statistic of Instruments | 236.0<br>(Pr>F=0) | 149.7<br>(Pr>F=0) |

Notes to Table 5: Sample includes students assigned to summer school who passed math but failed reading in June and who scored between 1 GEs below and 0.5 GEs above the reading cutoff in August. These results correspond to the first-stage estimates when reading scores two years later is the variable of interest. The exact results vary depending on subject and timing. August reading and math measures along with additional ability and demographic controls are included in both the first and second-stages.

**Table 6**
**The Effect of Grade Retention on Student Achievement**

| Dependent Variables | Specification | | |
|---|---|---|---|
| | OLS | IV | IV |
| | (1) | (2) | (3) |
| **Third Grade** | | | |
| Reading | | | |
| One year (n=8,120) | 0.085 | 0.162 | 0.174 |
| | (0.017) | (0.052) | (0.050) |
| Two years (n=7,623) | 0.012 | 0.026 | 0.035 |
| | (0.017) | (0.053) | (0.051) |
| Math | | | |
| One year (n=8,111) | 0.096 | 0.199 | 0.227 |
| | (0.019) | (0.058) | (0.044) |
| Two years (n=7,629) | 0.022 | 0.081 | 0.091 |
| | (0.019) | (0.059) | (0.045) |
| **Sixth Grade** | | | |
| Reading | | | |
| One year (n=5,018) | -0.137 | -0.077 | -0.064 |
| | (0.019) | (0.058) | (0.056) |
| Two years (n=4,552) | -0.176 | -0.160 | -0.154 |
| | (0.023) | (0.067) | (0.065) |
| Math | | | |
| One year (n=5,005) | -0.018 | -0.019 | 0.057 |
| | (0.020) | (0.060) | (0.040) |
| Two years (n=4,557) | -0.105 | -0.097 | -0.046 |
| | (0.021) | (0.063) | (0.047) |
| Additional Performance and Demographic Covariates | No | No | Yes |

Notes for Table 6: The sample includes third and sixth grade students from 1997 to 1999 whose June math score was above the cutoff, whose June reading score was below the cutoff and whose August reading score was between one grade equivalent below and 0.5 grade equivalents above the cutoff. Year fixed effects are included in each model. The additional performance and demographic covariates are the same as those listed in the notes to Table 2.

**Table 7**
**The Sensitivity of the Two-Year Grade Retention Estimates to Sample and Specification Choices**

| Specification | Third Grade | | Sixth Grade | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| Baseline estimates | 0.035<br>(0.051) | 0.091<br>(0.045) | -0.154<br>(0.065) | -0.046<br>(0.047) |
| Including second order polynomials in prior achievement | 0.048<br>(0.053) | 0.109<br>(0.047) | -0.162<br>(0.065) | -0.055<br>(0.048) |
| Including third order polynomials in prior achievement | 0.025<br>(0.079) | 0.108<br>(0.070) | -0.134<br>(0.123) | -0.059<br>(0.089) |
| Broader range of students | 0.001<br>(0.035) | 0.099<br>(0.032) | -0.139<br>(0.043) | -0.056<br>(0.032) |
| Broader range of students with second and third order polynomials in prior achievement | 0.073<br>(0.054) | 0.121<br>(0.049) | -0.188<br>(0.072) | -0.050<br>(0.053) |
| Narrower range of students | 0.015<br>(0.071) | 0.150<br>(0.063) | -0.183<br>(0.100) | -0.059<br>(0.073) |
| Alternative sample (students who passed the promotional cutoff in reading but not math in June) | -0.072<br>(0.105) | -0.145<br>(0.087) | -0.228<br>(0.090) | -0.147<br>(0.068) |

Notes to Table 7: Each cell contains an estimate from a separate 2SLS regression that controls for all of the past performance and demographic characteristics listed in the notes to Table 2. The baseline sample includes children who passed the promotional cutoff in reading but not math in June. The baseline range of data includes students who scored between 1.0 GEs below and 0.5 GEs above the reading cutoff in August. The broader range of data used in rows four and five includes students who scored between 1.5 GEs below and 1.0 GEs above the reading cutoff in August. The narrower range of data used in row six includes students who scored between 0.5 GEs below and 0.3 GEs above the reading cutoff in August.

**Table 8**
**The Heterogeneity of Two-Year Grade Retention Effects**

| Subgroup | Third Grade | | Sixth Grade | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| *Year* | | | | |
| 1997 Cohort | -0.134 | 0.053 | -0.388 | -0.234 |
| | (0.126) | (.109) | (0.140) | (0.112) |
| 1998 Cohort | 0.085 | 0.052 | -0.135 | -0.036 |
| | (0.065) | (.059) | (0.103) | (0.075) |
| 1999 Cohort | 0.060 | 0.184 | 0.001 | 0.013 |
| | (0.081) | (0.072) | (0.092) | (0.064) |
| *Prior Achievement*[a] | | | | |
| Bottom Quartile | 0.051 | -0.036 | -0.366 | -0.199 |
| | (0.102) | (0.087) | (0.147) | (0.105) |
| 2nd Quartile | -0.053 | 0.071 | 0.064 | 0.110 |
| | (0.090) | (0.079) | (0.120) | (0.087) |
| 3rd Quartile | 0.110 | 0.091 | -0.082 | 0.054 |
| | (0.118) | (0.109) | (0.121) | (0.091) |
| Top Quartile | 0.021 | 0.210 | -0.303 | -0.189 |
| | (0.119) | (0.109) | (0.138) | (0.101) |
| ***Race, Gender & SES*** | | | | |
| Black | 0.085 | 0.070 | -0.158 | -0.050 |
| | (.058) | (0.052) | (0.088) | (0.064) |
| Hispanic | -0.105 | 0.191 | -0.180 | -0.027 |
| | (0.110) | (0.097) | (0.097) | (0.069) |
| White/Other | -0.289 | -0.060 | -0.342 | -0.113 |
| | (0.203) | (0.161) | (0.312) | (0.219) |
| Male | 0.096 | 0.060 | -0.277 | -0.126 |
| | (0.069) | (0.062) | (0.090) | (0.064) |
| Female | -0.055 | 0.119 | -0.004 | 0.033 |
| | (0.075) | (0.066) | (0.094) | (0.070) |
| Free Lunch | 0.030 | 0.083 | -0.161 | -0.038 |
| | (0.052) | (0.046) | (0.066) | (0.048) |
| No Free Lunch | 0.134 | 0.213 | -0.188 | -0.140 |
| | (0.239) | (0.213) | (0.252) | (0.180) |

Notes for Table 8: Each cell includes an estimate from a separate 2SLS regression that controls for year fixed effects and all of the additional performance and demographic variables listed in the notes to Table 2. [a] Prior achievement is measured as the average math and reading score in second or fifth grade. Quartiles are determined on the basis of students in this sample. A small number of students who were missing second or fifth grade test scores are excluded from this categorization.

**Table 9**
**The Two-Year Effect of Summer School on Student Achievement**

| | Plausible Lower Bound | Estimated Effect (% of annual learning gain) | Plausible Upper Bound |
|---|---|---|---|
| **Third Grade** | | | |
| Reading | -0.009 | 0.053 (12.5%) | 0.116 |
| Math | 0.010 | 0.072 (11.7%) | 0.135 |
| **Sixth Grade** | | | |
| Reading | -0.019 | 0.031 (5.5%) | 0.081 |
| Math | -0.022 | 0.028 (5.8%) | 0.078 |

Notes for Table 9: The estimated summer school treatment effects are computed using the third difference two-year net effect estimates from Table 3 and the retention treatment effects from Table 6. We compute the treatment effect of the summer bridge program by removing that portion of the effect that could be caused by retention. For third and sixth grade, we compute lower and upper estimates by adding and subtracting 0.25 logits to point estimates of the retention treatment effects.

BEST COPY AVAILABLE